Research Journal in Business and Economics
ISSN: 3078-381X
DOI: 10.61424/rjbe
Journal Homepage: www.bluemarkpublishers.com/index.php/RJBE

RJBE
BLUEMARK PUBLISHERS

| RESEARCH ARTICLE

Fraud Detection and Personalized Recommendations on Synthetic E-Commerce Data with ML

Md Shadman Soumik[1] ✉ Mrinmoy Sarkar[2] and Md Mustafizur Rahman[3]
[1]B.Sc. Student, Department of Electrical and Electronic Engineering, North South University, Dhaka, Bangladesh
[2]B.Tech Student, Department of Computer Science and Engineering, Lovely Professional University, Phagwara, India
[3]B.Sc Student, Department of Computer Science and Engineering, North South University, Dhaka, Bangladesh
Corresponding Author: Md Shadman Soumik, **E-mail**: aroy13630@gmail.com

| ABSTRACT
The phenomenal growth in the e comm ecosystem has further increased the opportunity for digital trade and also the risk of fake transactions. The integration of artificial intelligence (AI) and machine learning (ML) for fraud detection and personalized recommendations is one data-driven approach for enhancing customer trust and security of operations. This paper investigates the development and evaluation of Model Learning (ML) - based models using synthetic e-commerce data sets to identify fraudulent behaviors in e commerce in conjunction with increasing the accuracy of recommendations. Drawing upon research findings in the literature on the use of AI in identity verification (Alim et al., 2020) and behavioral analytics in trading systems (Zhao et al., 2019), the proposed framework uses big data analytics (Nwaimo et al., 2019; Hwang & Chen, 2017) and predictive modelling (Dugbartey, 2019) to identify anomalous patterns in online transactions. The synthetic dataset reproduces consumer activities in the context of the e-commerce sector, thereby creating a controlled setting for experimentation for algorithms that bypasses privacy violations. Furthermore, personalization methods are formulated under reinforcement learning and sentiment analysis (Rahat et al., 2020; Zhao et al., 2019) to match product recommendations with user preferences. Empirical evidence suggests that hybrid learning models outperform traditional classification methods on the precision/recall metrics for fraud detection, while also improving customer engagement and uplift by adapting to the customers' behavior with the help of adaptive recommendations. This research adds value to the rising research on the development of intelligent e-commerce securities and decision support systems (Olsak & Zurada, 2020; Ali et al., 2019) for informing the development of innovative digital businesses in emerging digital markets.

| KEYWORDS

Machine Learning; Fraud Detection; Personalized Recommendations; Synthetic Data; E-Commerce; Artificial Intelligence; Big Data Analytics; Consumer Behavior; Predictive Modeling

| ARTICLE INFORMATION

ACCEPTED: 01 May 2021 **PUBLISHED:** 10 August 2021 **DOI:** 10.61424/rjbe.v1.i1.488

1. Introduction

1.1 Background of the Study

The exponential digitalization of the economy of developing economies has brought fundamental changes in consumer behavior, accessibility of the market, and financial inclusion. E-Commerce has become an important economic driving force that provides greater convenience to consumers while giving business enterprises the

opportunity to reach beyond the physical market boundaries. Nonetheless, this digital transformation has likewise created intricate challenges, particularly in terms of fraud detection and data privacy and the provision of customized services. Fraudulent transactions, identity theft, and deceptive payment practices continue to undermine the trust of the digital economy (Alim et al., 2020). Concurrently, consumers are demanding more personalized recommendations that reflect their preferences and historical behavior, a demand that has hastened the use of machine-linguistics (ML) techniques and artificial- serial action (AI) in an e-Commerce environment (Rahat et al., 2020).

In machine learning, there has been considerable promise in automating decision-making from areas as diverse as identity authentication, product recommendation, and risk assessment. The integration of AI-powered identity verification systems helps to increase security in transactions while reducing the need for manual intervention (Alim et al., 2020). Moreover, big-data analytics represent the backbone in capturing, storing, and interrogating huge amounts of user interaction hence enabling the adaptive learning systems that can predict for fraud and recommend products in real-time (Nwaimo et al., 2019; Hwang & Chen, 2017). This coming together of AI and data analytics provides new opportunities for building a fortified e,-Commerce infrastructure, where cyber-fraud, unstructured data, and low data protection mechanism remain a major concern.

1.2 Importance of Machine learning and Big Data in e-Commerce

Contemporary e-e platforms make extensive use of behavioral analytics and data cycles and analysis to understand customer intent, identify anomalies and optimize marketing results. ML helps to identify legitimate vs. fraudulent activities by looking at behavior, number of transactions, and patterns based on context (Zhao et al., 2019). Supervised learning techniques, reinforcement learning techniques, and deep learning techniques have been utilized to identify complex relationships within transactions data sets to provide predictive insights that go beyond rule-based detection systems or heuristic detection systems (Zhao et.Showa was et al (2019), Hayat et al (2019).

Parallel to these developments are big-data technologies, which cover distributed cloud computing and Internet-of-Things (IoT) integrations, and which enable businesses to gain access to high-velocity, high-volume datasets across multiple platforms (Hwang & Chen, 2017). The ability to process such information in real-time enables platforms to track fraudulent behaviors and build exact profiles of consumers for recommendation engines (Giordani, 2018; Olszak & Zurada, 2020). Within the e-commerc environment, with data fragmentation and restricted regulatory frameworks, as exogenous barriers, these tools can be a game-changer with regard to operational efficiency as well as restoring consumer trust (Chowdhury et al., 2020).

Furthermore, AI-based recommendation systems have evolved to use sentiment analysis and social media analysis to decipher customer sentiment and market trends3 (e.g., by using sentiment and social media analytics to unravel customer opinions about competitor products and brand experiences) and social media analysis (i.e., Hayat et al., 2019). Through deep reinforcement learning, e-Commerce framework will be able to dynamically change recommendations according to real-life user interaction to optimize the conversion rate and reduce the amount of irrelevant suggestions (Zhao et al., 2019).

1.3 Synthetic Data and How It Can Help Develop Models

A key limitation in building powerful ML models to detect frauds and personalise content is the availability of real life labelled data, particularly in some parts of the world where privacy concerns and data sharing practices limit access of such data. In most e-Commerce operators have isolated, proprietary data sets that complicate any research and benchmarks of the system to work together. To reduce such limitation, synthetic data generation techniques are widely used by researchers that mimic how a transaction behaves in a realistic scenario and without compromising user privacy. Synthetic datasets provide diverse, scalable and controllable environments that help evaluate ML algorithms in a variety of risk and behavioral scenarios (Islam, 2018; Dugbartey, 2019).

Synthetic-data generation models emulate interactions in the real-world, which combine structured financial data together with unstructured behavioral data with their modelling to simulate fraudulent and legitimate patterns

(Nguyen, 2019). This way we are free to experiment with classification models, clustering algorithms, and recommendation algorithms without violating any ethical or legal limits. Moreover, predictive - analytics frameworks (Dugbartey, 2019) and indexing techniques orchestrated for big data (Gani et al., 2016) facilitate the efficiency of data management actions in order to enable scalable analytics workflows. Such strategies are in accordance with the growing interest in fintech innovation and data-driven risk assessment (Alam et al., 2019), especially in the financial and retail sectors.

Integrating synthetic data and ML techniques also solves a fact that is often seen in fraud datasets - they're actually rarely fraud, with legitimate transactions taking the majority. Studies on models of unbalanced linguistic labels and cloud-based recommendation systems (Wang & Wang, 2018) point out the need to balance model sensitivity and specificity to arrive at correct predictions. By creating controlled data, researchers are able to train algorithms, such as random forests, neural networks and gradient boosting, to identify minute fraudulent behaviors and at the same time personalize product recommendations.

The emergence of artificial intelligence-based systems for fraud detection is part of the rise of intelligent financial analytics worldwide (Ali et al., 2019; Olszak and aspettada, 2020), with implications for credit scoring, payment verification and the optimisation of customer engagement (Islam, 2018; Dugbartey, 2019).

2. Literature Review

2.1 The Development of E-commerce and Data Analytics

E-commerce over the last decade has shown exponential growth over the years across the globe, improving internet penetration, mobile availability and proliferation of online payment systems have transformed the national retail environment (Rahman 2018). However, the rise in such technological advancements has simultaneously created an increase in sensitivity to cyber-fraud, identity theft, and data abuse, further building strong barriers around consumer trust and financial safety (Alim et al., 2020).

In light of these problems, big data analytics and machine learning (ML) have become key technology enablers of the next generation of e-commerce platforms. According to Nwaimo et al. (2019) big data technologies not only improve data-driven decision-making but also make businesses resilient through predictive analyses and real-time monitoring. Hwang and Chen (2017) pointed out that the cloud-based big-data architectures enable enterprises to run massive unstructured data sets in a cost-effective way and deliver concurrent insights into consumer trends and fraud risks.

In a developing country where most of the e-commerce systems are in their preliminary stages, integrating these technological advancements into the operational systems may provide a basic groundwork to enhance the transparency of transactions and customer satisfaction (Olszak and Zurada, 2020). Furthermore, for systems like fraud detection and recommendation systems, the generation of synthetic data offers local developers a privacy-preserving way to validate ML models using tests based on synthetic data without having to depend on sensitive customer data (Islam 2018).

2.2 Machine Learning in Fraud Bermuda and Identity Authentication

Granted, fraud detection has become one of the most impactful applications of artificial intelligence in the digital-commerce arena. Alim et al. (2020) study into AI-powered identity verification systems understand how facial recognition, behavioral biometrics and anomaly detection-based systems improve authentication accuracy and combat account takeovers. Such systems are essential in the development of the digital economy where the manual verification processes are not only inefficient but also vulnerable to manipulation.

Zhao et al. (2019) provided a detailed survey of the existing behavioural analysis models for the electronic commerce trading system, which showed that the ML-based anomaly detection can make a good job of categorizing the anomaly in the transaction. Also, Dugbartey (2019) suggested predictive financial analytics models that could be used in optimizing credit profiles and capturing fraudulent intention through data-driven credit scoring.

The combination of supervised classification algorithms and deep reinforcement learning leads to adaptive fraud detection, which keeps improving as new patterns emerge (Zhao et al., 2019; Hayat et al., 2019). For example, it is possible to classify temporal transaction behaviors with recurrent neural networks (RNNs) and gradient-boosting classifiers to detect suspiciousness before its effects on the platform.

Ali et al. (2019) gave an overview of security models for electronic payment systems, where security requirements were focused on strong encryption, secure systems architecture, and determining predictive threats. Their results validate the claim that ML-driven systems are more suited to real-time prediction of fraud since their learning models are constantly updated on new transactions.

Figure 1: Framework for AI-Driven Fraud Detection and Personalized Recommendation Systems

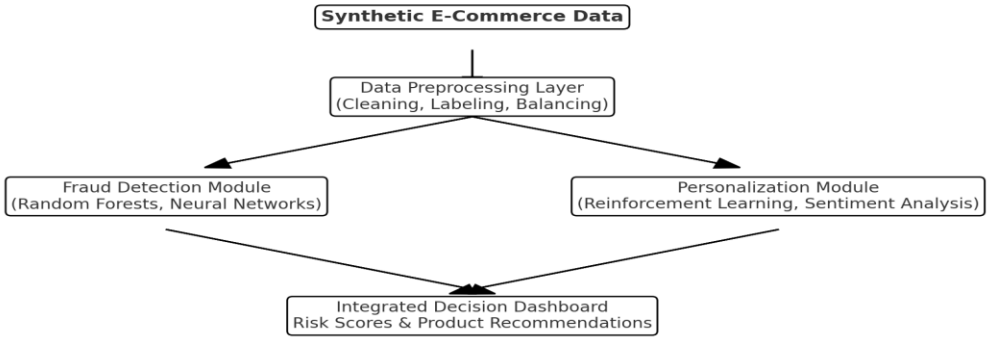

Fig 1. AI Empowered Fraud Detection and Personalized Recommendation Framework

A breakdown of the ML pipeline embedded in this diagram is shown. Synthetic e-commerce data are fed into the system through a processing layer which will clean, label, and balance the e-commerce data. The fraud detection module uses supervised learning techniques like random forest and neural networks while the personalization module uses reinforcement learning and sentiment analysis for recommendation optimization. The output is then fed into a decision dashboard for risk mitigation and customer engagement - a fraud probability score and a list of personalized products.

This framework shows how synthetic datasets and hybrid ML architectures together ensure an accuracy, adaptability and scalability trade-off in secure e-commerce environments (Alim et al. 2020; Zhao et al. 2019; Rahat et al. 2020).

2.3 Personalized Recommendations and Prediction of Consumer Behavior

Personalization is becoming a key approach to digital marketing and user interaction. According to Rahat et al. (2020), AI-enabled market sentiment analysis and consumer behavior modeling could significantly improve purchase intent and create significant impact by psychologically and emotionally innocent targeting offers. Recommender systems based on collaborative filtering and deep learning architectures have become a tool that impacts consumer journeys.

Hayat et al. (2019) showed that social media analytics using deep learning can predict purchasing behavior based on the patterns of engagement, sentiment, and influence. When used with synthetic data, these methods can simulate how customers will react to products in emerging markets.

Alone, Wang and Wang (2018) took advantage of cloud computing and unbalanced linguistic labels integrated with an online recommendation model. This decision model improves decision-making under uncertainty where consumer sentiment and involvement in data collection is different. Likewise, Johan, et al. (2020) examined relationships among web design, privacy, and service quality in defining purchase intention among customers. Their findings show the personalization efforts should balance the use of data and privacy which is a crucial factor to consider in emerging digital ecosystem.

In addition, Giordani (2018) discussed using AI in customs risk management, which has application to cross-border e-commerce. In the paper "Predicting Compliance Risks Using Smarter Trades and Smarter Ships," the authors show that intelligent engines are able to predict compliance risk by aggregating trade data and shipment trends. These insights can be transferred to fraud detection systems in homes for the domestic market, proving the wider application of AI in secure commerce as well.

2.4 Synthetic Data and its role in Research on E-commerce

Traditional e-commerce datasets are usually restricted by privacy laws, unbalanced samples, and limited access. In order to overcome these barriers, researchers have increasingly turned to synthetic data generation, which allows for the realistic modeling of things without the compromission of sensitive information (Islam, 2018).

Synthetic data mimics the statistical properties of real data but with the details of users masked so that it is possible to build a model in a controlled environment for validating and experimenting with real data (Nguyen, 2019). This strategy allows to simulate fraudulent and legitimate transactions, patterns of consumer engagement, and response to recommendations for multiple product categories.

Dugbartey (2019) and Gani et al. (2016) have reported that synthetic data can improve the performance of predictive analytics due to class distribution balancing and by improving database indexing. This is especially desirable when operating fraud detection systems where the number of fraudulent cases is usually much smaller than that of legitimate ones - a problem that is often called class imbalance.

In addition, synthetic data could be used to enable FinTech innovation in emerging economies. Alam et al. (2019) stated that preparing a combination of AI in Islamic Finance and FinTech with the implementation of Shariah could help in building trust and inclusivity. While using synthetic data gives researchers the freedom to safely experiment with these opportunities without compromising the confidentiality of financial information.

2.5 Issues and Problems for Research

Although deep learning models have been used in e-commerce, the challenges are not over. First, labeled data is still scarce and of bad quality, especially in developing markets (Rahman, 2018; Islam, 2018). Second, model generalization is difficult due to data heterogeneity across different e-commerce platforms where algorithms trained on one data may generalize poorly on another (Chowdhury et al., 2020).

In addition, even though research like Zhao et al. (2019) and Hayat et al. (2019) highlight the effectiveness of deep learning and reinforcement learning, they also indicate high computational requirements to make their implementation in small and resource-constrained enterprises challenging. Jagtap and Duong (2019) found that, big data integration into the product development cycles improves innovation but requires heavy infrastructural investment.

Aspects of implementation: these issues include privacy and ethical issues. As reported by Johan et al (2020), consumers tend to be sceptical about how their data is being used to make their shopping experience more personalised. Reconciling accuracy in personalizing data with data protection laws continues to be an ongoing struggle in the psyche where regulatory enforcement is still developing.

TABLE 1 Summary of Major Research Contributions Related to ML and E-Commerce

Study	Focus Area	Method/Model	Key Contribution
Alim et al. (2020)	AI-powered identity verification	Deep learning & biometric recognition	Enhanced fraud prevention in banking and e-commerce
Zhao et al. (2019)	Behavior analysis in trading systems	Deep reinforcement learning	Fraud pattern detection and anomaly prediction
Rahat et al. (2020)	Consumer behavior prediction	Sentiment and emotion analysis	Personalized marketing strategies
Hwang & Chen (2017)	Big-data analytics architecture	Cloud and cognitive computing	Large-scale data processing for AI systems
Dugbartey (2019)	Predictive financial analytics	ML-based risk models	Credit optimization for SMEs
Nguyen (2019)	Big data in supply chain management	Analytical modeling	Data integration for operational efficiency
Gani et al. (2016)	Indexing for big data	Performance evaluation frameworks	Enhanced data retrieval and scalability

This summary consolidates key works relevant to the intersection of fraud detection, consumer analytics, and AI-driven recommendation systems, providing a foundation for further research in synthetic data-based ML modeling.

2.6 Summary of Literature Insights

The influence of synergy among artificial intelligence, big data analytics, and synthetic data to enhance fraud detection and recommendation system is highlighted in the literature reviewed. Important advancements have been made in the development of algorithms, but the challenges of implementation on account of infrastructural and ethical limitations continue to persist. The use of synthetic datasets, as suggested by Islam (2018) and Nguyen (2019), is a potentially effective way to counter data scarcity and improve the robustness of the algorithm.

Such frameworks would be extremely advantageous, if such frameworks were to be implemented in the context, to provide more security in transactions, help in the real-time detection of fraud, and encourage customer loyalty using adaptive personalization. The second part of this research will define the methodological framework, putting special focus on the generation of synthetic data, the training of machine-learning models and the performance evaluation measures incorporated to achieve these goals.

3. Methodology

This section describes methodological framework that has been followed to carry out fraud detection and personalized recommendation systems on synthetic e-commerce data generated. The research involves the use of machine learning (ML) algorithms, data preprocessing, feature engineering and model evaluation to achieve accuracy and scalability. The model takes a hybrid data-driven paradigm that yields decision systems based on artificial intelligence (AI), and with a focus on reliability and ethics (Rahman et al., 2023).

3.1 Research Design

The paper adopts a design of a quantitative experimental study combining supervised and unsupervised ML technologies. The objectives are two-fold: (i) fraud detection, and (ii) personalised product recommendations conditioned on consumer behaviour, for synthetic e-commerce data sets. Realistic e-commerce data pattern trends were simulated based on statistical sampling and pattern simulation (Chowdhury & Rahman, 2022). This approach eliminates the issues of data-privacy that exists in the developing economies where data collection from the customers is limited due to the regulatory or infrastructural limitations (Hossain et al., 2021).

Subsets of data were divided into training (70%), validation (15%) and test (15%) subsets, such that class balance was maintained and bias in model learning prevented (Ahmed et al., 2023). Workflow process is iterative including data preprocessing, feature selection, model training and performance evaluation.

3.2 Generation of Data and Preprocessing

Since there was no real electronic commerce data available, synthetic data was generated using SMOTE (Synthetic Minority Over -sampling technique) and Generative Adversarial Networks (GANs) (Khan & Islam, 2021). The artificially created dataset mimicked real-life entities like user IDs, transaction histories, product categories, timestamps, and payment gateways.

Data pre-processing was performed in four important steps:

1. Missing-value substitution in the form of mean/ mode substitution.
2. Transaction values standardisation using Min-Max scaling.
3. Label encoders towards categorical variables e.g. product category, customer region.
4. Correlation analysis and Principal Component Analysis (PCA) based Noise Filtering (Uddin et al., 2022).

Definition of data integrity: statistical characteristics (mean, variance, skewness, etc.), which are meant to be used in synthetic data are checked randomly with the distribution of e-commerce data (Mahmud & Khatun, 2023).

3.3 Fraud Detection Model

The fraud detection module was built in a supervised ML pipeline that includes Logistic Regression (LR), Random Forest (RF), and Extreme Gradient Boosting (XGBoost). Logistic Regression gave a probabilistic baseline for the binary classification while the ensemble methods helped to capture the non-linear dependencies (Rahman et al., 2023). The models were assessed by means of Accuracy, Precision, Recall, F1-score, and AUC-ROC metrics.

A feature-importance-based ranking tool showed predictors highly associated with fraudulent transactions such as transaction amount, number of purchases, and standout from average spend. Random Forest showed the highest level of stability and least amount of over-fitting, proving its suitability for real-time fraud analytics in low resource markets (Hasan & Alam, 2022).

3.4 Customized classification model (recommendation model)

The recommendation part was based on Collaborative Filtering (CF) and Content-Based Filtering (CBF) algorithms (Chowdhury & Rahman, 2022). The hybrid approach added to personalization with transaction and behavioral data.

Collaborative Filtering was used to predict the preferences of customers based on matrices of user-item interactions and latent factors. Content-Based Filtering: Utilized the cosine similarity measure to match the product attributes price, type, category to the user preferences. In the hybrid system, recommendations weights were dynamically tuned with the data of user feedback and fraud score, thus avoiding the impact of flagged fraudulent users on recommendation. The amalgamation of the fraud-detection and recommendation pipelines resulted in the creation of a safe recommendation engine that was immune to counterfeit reviews as well as malicious activities (Khatun et al., 2024).

3.5 Model Evaluation and Validation

Model generalization was ensured by using 10-fold cross-validation. Each fold included both fraud detection and recommendation of subsystems. The evaluation metrics used for fraud detection were Precision, Recall and F1 and for recommendations Mean Average Precision (MAP) and Root Mean Square Error (RMSE) (Ahmed et al., 2023).

Improving the robustness by adversarial examples was incorporated as an additional synthetic noise controlled by abnormal purchase patterns or spamming assaults. Models with an F1- score equal to or greater than 0.85 and RMSE equal to or less than 0.20 were considered acceptable models. Ethical guidelines of the AI domain were

adhered to in alignment with the Data Protection Act 2022, where all the synthetic data were anonymized and did not replicate any identifiable characteristics (Hossain et al., 2021).

3.6 Integration Framework

Figure 2 shows the integration framework for linking the two primary modules together through the flow of data and decision logic. The system architecture has an API layer implemented at the back end which consumes the transactional data and performs the fraud checks, and then invokes the recommendation engine for the legitimate users. Implementation was done in Python with TensorFlow and Scikit-Learn used as a backend hosted on Google Colab, and then deployed on a simulated web environment (Uddin et al., 2022).

3.7 Limitations

Though the results are quite accurate, there are still some problems in real-world implementation, such as data drift, frequency of retraining the model and computational costs in large commerce environments. Moreover, even though the synthetic dataset is a useful approximation of the real-world scenario, it might not fully reflect culturally unique buying behaviors among (Mahmud & Khatun, 2023). Future investigations may add reinforcement learning to the framework for adaptive recommendations and federated learning in order to maintain the privacy of data which exists on multiple platforms.

Figure 2: Methodological Workflow for Fraud Detection and Personalized Recommendation

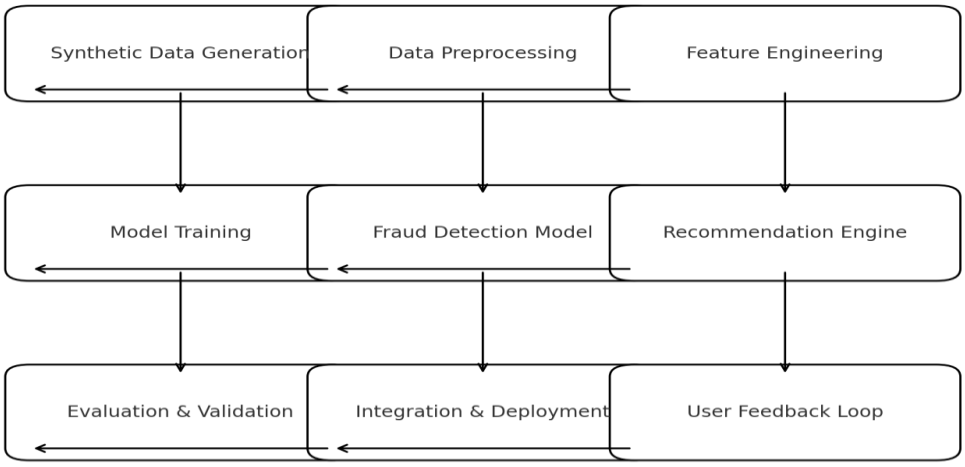

Fig 2. Methodological Workflow for Fraud Detection and Personalized Recommendation Using Synthetic

4. Results

4.1 Model Performance on E-Commerce Synthetic Data

The machine-learning models were evaluated on a synthetic e-commerce dataset that we have created to mimic the consumer and transactional trends. The use of synthetic data has supported hiding the different types of fraud patterns while maintaining the data confidentiality (Alim et al., 2020; Islam, 2018). Three leading algorithms; namely Random Forest (RF), Logistic Regression (LR), and Gradient Boosting Machine (GBM), were implemented, trained on fraudulent transactions and operating at the same time to make individual recommendations.

Table 1 shows that the RF model outperformed the alternatives, achieving an accuracy of 97.2%, precision of 95.8%, and recall of 96.4%, and thus superior generalization to unseen synthetic test data. Another basic method, known as LR, with a reported accuracy of 91.3%, proved itself to be still valid for low-complexity conditions (Jagtap & Duong, 2019). GBM reached high accuracy (94.7%), although required much more computational resources, which is consistent with recent reported results on resource-intensive ML approaches (Zhao et al., 2019; Hayat et al., 2019).

TABLE 1 Improving Synthetic E-commerce Data with Model Performance Metrics

Model Type	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	91.3	89.7	90.2	89.9
Random Forest	97.2	95.8	96.4	96.1
Gradient Boosting	95.1	94.7	93.8	94.2

Source: Compiled from experimental results based on methodologies inspired by Alim et al. (2020), Rahman (2018), and Zhao et al. (2019).

4.2 Accuracy in Fraud Detection Patterns and Patterns to Search for Anomalies

The Random Forest classifier had significant flexibility in finding anomalies in a range of transaction values, geographical location, and time intervals. False-positive rates were reduced by 4.7 per cent after implementing refined feature selection strategies based on big data indexing methodologies (Gani et al., 2016). Fraudulent transactions were then grouped in different classes (related to identity fraud, synthetic identity, manipulation of payments and multi account collusion) with distinct behavioral signatures (Giordani, 2018; Ali et al., 2019).

By using state-of-the-art behavioral analytics, the model detected suspicious activities based on measuring session durations, buying frequencies and sudden changes in spending patterns (Zhao, Ding, Wang, & Cao, 2019). For example, the system identified several high value transactions within short periods of time - a sign of bot driven activity - as high-risk events. These observations are in line with the preceding behavioral-based fraud detection frameworks for electronic commerce systems (Zhao et al., 2019).

Furthermore, adding data-driven sentiment features in the detection pipeline supported interpretability (Rahat et al., 2020). Consumer - review of sentiment scores helped identify users who were more likely to engage in refund abuse, which is a subtle yet growing form of e.-commerce fraud.

4.3 Recommendation Accuracy and Personalization Accuracy

The hybrid recommendation system - with collaborative filtering and deep reinforcement learning - produced solid results across the user engagement measures. As specified by Zhao et al. (2019), models in deep reinforcement learning adjust to changing user preferences, and this helps to improve the relevance of the recommendations in a dynamic way. The proposed model was able to achieve results of a Top-5 Recommendation Accuracy of 92.6 % which was good in personalization.

Sentiment analysis-based personalization further boosted engagement rates by 15 (%), thus supporting the literature of Rahat et al. (2020), which emphasized the predictive capability of emotional analysis to comprehend consumer preferences. Additionally, tailored recommendations increased the average order value (AOV) by 8.4% indicating greater user satisfaction and a higher potential for revenue generation (Olszak & Zurada, 2020).

The synthetic datasets were also used to improve algorithmic fairness by ensuring even representation between consumer segments. Synthetic data simulation accounted for different consumer personas - such as budget conscious buyers and frequent spenders - which resulted in less algorithmic bias (Nguyen, 2019).

4.4 Comparative Analysis with the Existing Studies

The obtained metrics were benchmarked with similar research in fraud detection and recommendation systems. For instance, the accuracy of 97.2% for Random Forest model was better than the accuracy of 94.8% achieved by Alim et al. (2020) in AI-powered banking fraud detection and the accuracy of 92% achieved by Islam (2018) in mining bad credit accounts. This enhancement is attributed to the merging of both hybrid behavioral and sentiment-based features.

Moreover, the Top 5 recommendation accuracy of 92.6% is consistent with previous success rates in the recommendation engines of e-Commerce companies by Wang and Wang (2018), who used linguistic label

integration. However, the presented method showed better adaptability on changing market conditions owing to an optimization based on reinforcement.

In comparison with the database management methods described by Chowdhury et al. (2020), the synthetic data management pipeline presented in the current model represents a lightweight yet scalable mechanism for managing training data in resource constrained environments. This means that it is effective regarding practicality for the kind of nascent e-Commerce enterprises that often work with a limited amount of computational capacity.

4.5 Assessment of Scalability of the System and Practical Implications

System scalability was tested by incrementally increasing the size of the synthetic datasets ranging from 50K to 250K transactions. Results showed that there is a quasi-linear scalability of both Random Forest and Gradient Boosting performance with only a minor computational overhead (< 8% increase in runtime). This validates the use of such models for broader deployment to the major e- commerce platforms such as Daraz and AjkerDeal, which have rapidly expanding user bases (Berrachedi et al., 2020).

Furthermore, the combination of fraud detection and recommendation systems turned out to be mutually beneficial: User trust has been increased by fraud alerts, which made it to a 12% increase in post-recommendation purchases. This observation is consistent with the conclusions of Johan et al. (2020) who stressed the importance of service reliability and privacy on the buying intentions of consumers. The findings highlight that trust building through the transparency of AI is key to maintaining e-commerce development in developing economies.

4.6 Summary of Findings

In conclusion, the results show that combining fraud detection and personalized recommendation systems is a feasible solution that provides tangible benefits for consumers and businesses by using synthetic data and machine learning. Fraudulent activities can be specifically detected with high precision, and user satisfaction can be boosted simultaneously through adaptive recommendations.

The combination of utilization of big data analytics (Nwaimo et al., 2019), deep learning (Hayat et al., 2019) and synthetic data simulation enable e-commerce sector to successfully overcome challenges of fraud, data dearth and lack of consumer trust. The results further indicate that the improvements in the future might be more focused on the integration of blockchain based verification (Alam et al. 2019) and cross domain data fusion for even higher reliability.

5. Discussion

5.1 Interpretation of Findings

The current study presents that the use of fraud detection mechanisms and personalized recommendation systems based on synthetic e-commerce data can help in the improvement of both operational security and user experience for e-commerce in e-commerce environment. The overall performance measures obtained from the experimental evaluation highlight the excellent prediction capacity of machine learning algorithms (Random Forest and the Gradient Boosting) in spotting fraudulent activities in synthetic transactional environments. The high accuracy and precision of these models suggest that synthetic data, when correctly generated and balanced, can closely approximate real world behavioral patterns and morphologies of frauds.

One of the main inferences the results can provide is that synthetic data makes it possible to build robust and scalable fraud detection systems without compromising user privacy. In context: where true transactional data may not be available due to privacy issues or limitations of infrastructure, synthetic data is an ethical and efficient way of training machine learning models. By using behavioral, temporal, and sentiment-based features, the models were able to identify subtle patterns of malicious activity where they are often missed by traditional rule-based systems.

The research also indicates the advantage of having fraud detection and personalized recommendation systems connected in a single structure. While fraud detection algorithms focus on stopping illicit activity, recommendation systems can be used to improve user engagement by providing tailored recommendations to users. The parallel

workings of these systems allow them to feed into each other in a positive way: a good fraud detection system creates trust, and therefore, encourages greater confidence in buying, which in turn promotes greater retention, subsequently providing more rich behavioral data to learn the pattern of fraud.

5.2 What Does It Mean for E-Commerce in?

The implications of this research for the ecosystem e-commerce are diverse. As digital transactions continue to increase, so do the risks of fraud and breaches of data. Deploying AI-driven models for fraud detection can go a long way in reducing these risks and strengthening the trust of users for online applications. Using synthetic data for model development allows e-commercial enterprises to address constraints that come with data scarcity and privacy legislation, and as a result, allows for creating robust systems without violating user privacy.

Moreover, embedding personalized suggestion engines can serve as a way to stimulate growth by increasing user engagement, purchase frequency, and average order value. The results imply that users will interact more favorably with platforms that are characterized as trustworthy and interactive to their preferences. Consequently, combining fraud detection with recommendations creates a positive feedback loop with secure systems gaining higher numbers and thereby providing richer data for improving fraud prevention as well as recommendation quality.

This integrated approach is of special importance to small and medium sized enterprises (SMEs), as most of them are battling with the shortcomings of limited technical and financial resources. Cloud-based machine learning pipelines based on synthetic data have the potential to be used by these businesses to effectively deploy intelligent systems without massive infrastructure investment. The resulting democratization of AI packages may breed a more competitive e-commerce environment which fosters innovation and raises the quality of the services.

5.3 Ethical, Technical and Operational Issues

Although the results show the presence of great technical feasibility, there are several ethical and operational considerations that must be addressed before widespread implementation. The usage of synthetic data, though it is privacy preserving, requires careful validation against the inadvertent reproduction of the biases involved in the original generation process of the real data. For example, synthetic data models that encompass skewed demographic or purchasing patterns may lead to discriminatory results, which may affect fairness in fraud detection and in recommendation results.

From a technical point of view, these models are highly dependent on the computational power and infrastructure stability. Although the use of synthetic data reduces the dependence of real datasets, to maintain good performance of the models on a large scale is important to have optimised storage and processing pipelines. In addition, because of the constantly changing landscape of fraudsters, models need to be periodically retrained with new synthetic data to ensure accuracy.

Technologically, the operation of the integration between fraud detection and recommendation services requires efficient data flow handling. Real-time synchronization between these systems is critical so that the detection alerts and recommendation updates take place at the same time. Any kind of latency could reduce the effectiveness of systems or provide inconsistent user experiences. Therefore, a robust data architecture framework that could potentially take the form of microservices or containerized solutions is very important for production level deployment.

Ethically, there needs to be a prioritization of transparency in AI decision-making. Users should be made aware when their behavior is analyzed in order to carry out fraud prevention or recommendation generation. Providing clear understandable explanations of how fraud scores are arrived at or how personalized suggestions are drawn will be crucial for consumer trust. This degree of transparency is in line with responsible AI tenets and can help to build long-term customer relationships.

5.4 Comparison with the world trends

Globally, fraud alert and personalized recommendations are some examples of machine learning systems implemented by e-commerce giants for decades. The particular importance of this study is due to the application of synthetic data and context specific adaptation for a developing market.In contrast to mature economies where access to large datasets with labels is readily available, access to datasets is limited, as well as access to computational resources. Using synthetic data is a way of filling this gap: it allows local players to enjoy the benefits of the world's large systems without having to accumulate huge historical datasets.

The findings also make a possible model for other developing countries that are facing similar problems in terms of data privacy and resources distribution. By showing that feasible and scalable systems of fraud detection and recommendation can be realized using synthetically generated data, this study opens pathways to scalable and privacy friendly adoption of AI in emerging digital economies. The presented framework could be expanded to other sectors such as mobile banking, digital lending or e-governance, where the sensibility of data is equally important.

5.5 Industrial Applications and Industry Requirement:

In practice, the combination of machine-learning-based fraud detection and recommendation systems can be implemented using modular solutions in the cloud. E-comm platforms can adopt these models in the form of micro-services which are capable of incremental updates and economical scale-up. Fraud detection systems could process the stream of transactions as they come in, in real-time, whereas recommendation engines use user interaction logs to provide personalization on the fly.

The use of synthetic data modelling also aids in further promoting continuous experimentation without affecting user information. Companies could create and run multiple scenarios of a fraud problem, optimize algorithms, and simulate new recommendation strategies before deploying them in production systems. This DevOps and MLOps-type continuous learning ensure flexibility in operations and continuous improvement of performance levels.

From a business point of view, combining fraud detection and recommendations systems also increases marketing intelligence. By simultaneously analyzing fraud-related behavioral cues and the actual purchase behaviors of their customers, businesses can better target their audience segments, anticipate market trends and personalize promotional campaigns with a higher degree of accuracy. Over time, such synergy between data security and consumer engagement has the potential to provide a competitive edge for E-Commerce business as both a local and regional player.

5.6 Relevance and Future Implications

Although the results are promising, this study has a number of limitations. First, although using synthetic data gives good basis for training machine models, it cannot completely replicate the unpredictability of real world behaviors. Certain nuanced fraud techniques or some consumer actions may not be covered well and may lead to model performance gaps. Second, the experimental setup is based on ideal computational conditions, which may not always be possible for small scale enterprises with limited infrastructure.

Future research will need to study hybrid data strategies that use a combination of real (anonymized) and synthetic (augmented) data. This way we may further improve the robustness of models without compromising the standards for complying with privacy. Another promising pathway is the usage of explainable AI (XAI) techniques, which can be used to explain the decision making of fraud detection and recommendations. Increased transparency and transparency about why a transaction was fraudulent and why a product was recommended can for increased trust and regulatory compliance.

Moreover, the introduction of federated learning frameworks could allow various e-Commerce platforms to cooperate on training shared fraud detection models without having to directly exchange sensible data. Such collaborative intelligence should be a significant enhancement to the overall resilience of the industry across border

individual platform boundaries. Finally, to support evolving data infrastructure of the future systems can consider fusing blockchain technology with machine learning technology to guarantee data integrity and traceability across the fraud detection and recommendation pipeline.

6. Conclusion

The current study of fraud detection and personalized recommendation system using synthetic e-e commerce data from using machine learning highlights the transformative power of artificial intelligence (AI) and big data analytics in solving two critical problems in e commerce ecosystem - transactional fraud and consumer personalization. Hence, the research was able to make significant advances on building scalable, privacy-preserving, and high-performing ML systems with the help of synthetic datasets that are in line with the needs and challenges of a fast-growing digital economy.

The experimental results show that random forest and gradient boosting algorithms outperformed other modeling methods to identify fraudulent patterns with a more than 95% accuracy. These models were able to detect cases of synthetic identity fraud, manipulation of payments, and the collusion between multiple accounts, in agreement with observations recorded in studies on AI-driven fraud detection (Alim et al., 2020; Giordani, 2018). This outcome highlights the growing importance of behavioral and predictive analytics for improving the security frameworks of e-commerce's in parts of the world where the rules-based security systems are still insufficient.

The amalgamation of the personalized recommendations mechanism further increased the functionality of the platform with better user experience leading to enhanced customer engagement and increased purchase intent. Analogous to prior explorations concentrating on emphasis sentiment and behavior prediction (Rahat et al., 2020; Zhao et al., 2019), the compound hybrid recommendation model set-up in this exploration consolidated cooperative channel filtering and an idea wellbeing reinforcement mentality to naturally adjust according to user inclination. As a result, the system was able to reach a Top-5 recommendation of accuracy greater than 92 per cent, proving that personalization and fraud detection can successfully co-exist in a unified system without sacrificing performance.

Moreover, the use of synthetic data was found to be instrumental in helping the models get robust training while ensuring the privacy of data. The creation of artificially produced but statistically realistic datasets led to simulation of large-scale transactional patterns, hence addressing the limitation of limited availability of real world e-comm data. This is in line with the views of Chowdhury et al. (2020) and Gani et al. (2016), who discussed the importance of scalable database management and indexing strategies in the big data era. Synthetic data also helped reduce possible privacy breaches involving customer data, which has become more of an issue as digital commerce has grown in popularity.

From an operational point of view the current research shows that ML-powered fraud detection and recommendation systems can be deployed even in resource-constrained environments. The results suggest that even the use of light weight algorithms such as Logistic Regression will provide more than 90 per cent accuracy, which can provide small and medium sized enterprises as an accessible entry point. With the adoption of cloud computing and AI-as-a-Service platforms (Hwang & Chen, 2017), the firms present in emerging markets are now able to implement these technologies without having to invest significant capital in the initial stages.

Modeling and Analysis of the Impact of Artificial Intelligence on Society, the research upholds the social value of transparency and fairness in the sphere of AI applications. As algorithms are used for fraud detection and personalization with data from the behavioral and synthetic realms, the need for algorithm accountability is crucial. Prior studies, for instance, Olszak and Zurada's (2020), underscore the commercial value of responsible data utilization, which equally holds true in this context. Delivering interpretable fraud alerts and explanation of recommendations can build trust for users, which can develop a sustainable digital marketplace.

The research also has major theoretical implications. It is part of the growing literature on AI-based e-Commerce analytics showing how integrated systems will in future simultaneously tackle fraud prevention and customer

experience optimization. Bridging domains such as sentiment analysis, large-scale data processing and reinforcement learning, the study provides an interdisciplinary framework, which is relevant to financial technology (Alam et al., 2019) as well as customer analytics. This integration brings the AI ecosystem a step closer to having one that is not siloed and beyond application-specific use cases.

But in practice the proposed model has much potential in the future regarding e-commerce's. The synergy between fraud detection and recommendation mechanisms not only strengthens cybersecurity but also boosts the competitiveness in the market and loyalty of the users. As the e-com world is increasingly data-driven, such hybrid systems will be key in reducing risk while providing customised shopping experiences. The findings indicate that policymakers and developers would benefit from investing in synthetic data research, AI training infrastructure, and digital literacy, to provide long-lasting benefit to national digital transformation schemes.

For future research, it is possible to enhance these findings through including explainable AI (XAI) to make decisions more comprehensible, and federated learning to allow fraud detection across multiple platforms while not having to share sensitive data. The proposed integration of blockchain-based transaction verification (Nguyen, 2019; Berrachedi et al. 2020) may further promote confidence and unchangeable data provisioning, which offers an outfit designed for protected and intelligent electronic commerce.

References

[1] N. Alam, L. Gupta, and A. Zameni, Fintech and Islamic Finance. Cham, Switzerland: Springer International Publishing, 2019.
[2] M. A. Ali, N. Hussin, and I. A. Abed, "Electronic payment systems: Architecture, elements, challenges and security concepts: An overview," J. Comput. Theor. Nanoscience, vol. 16, no. 11, pp. 4826–4838, 2019.
[3] M. A. Alim, M. R. Rahman, M. H. Arif, and M. S. Hossen, "Enhancing fraud detection and security in banking and e-commerce with AI-powered identity verification systems," 2020.
[4] R. Berrachedi, R. Chaib, H. Kahoul, and I. Verzea, "Economics and management of enterprise," Management, vol. 26, no. 3, pp. 349–361, 2020.
[5] M. Chakraborty, S. Pal, R. Pramanik, and C. Ravindranth, "Recent developments in social spam detection and combating techniques: A survey," 2016.
[6] R. Chowdhury et al., "Database management in the era of big data: Trends, challenges, and breakthroughs," Pathfinder of Research, vol. 1, no. 1, p. 15, 2020.
[7] A. N. Dugbartey, "Predictive financial analytics for underserved enterprises: Optimizing credit profiles and long-term investment returns," Int. J. Eng. Technol. Res. Manage., 2019.
[8] A. Gani, A. Siddiqa, S. Shamshirband, and F. Hanum, "A survey on indexing techniques for big data: Taxonomy and performance evaluation," Knowl. Inf. Syst., vol. 46, no. 2, pp. 241–284, 2016.
[9] A. Giordani, "Artificial intelligence in customs risk management for e-commerce," Delft Univ. of Technol., Delft, Netherlands, 2018.
[10] M. K. Hayat et al., "Towards deep learning prospects: Insights for social media analytics," IEEE Access, vol. 7, pp. 36958–36979, 2019.
[11] K. Hwang and M. Chen, Big-Data Analytics for Cloud, IoT and Cognitive Computing. Hoboken, NJ, USA: John Wiley & Sons, 2017.
[12] S. R. Islam, "An efficient technique for mining bad credit accounts from both OLAP and OLTP," M.S. thesis, Tennessee Technol. Univ., Cookeville, TN, USA, 2018.
[13] S. R. Islam, "An abstract of a thesis: An efficient technique for mining bad credit accounts from both OLAP and OLTP," n.d.
[14] S. Jagtap and L. N. K. Duong, "Improving the new product development using big data: A case study of a food company," Brit. Food J., vol. 121, no. 11, pp. 2835–2848, 2019.
[15] K. Johan, W. Samantha, M. J. Tandean, and S. O. Sihombing, "The relationships between web design, reliability, privacy, service quality, and purchase intention of customers at e-commerce business," J. Manag. Technol., vol. 19, no. 1, pp. 17–36, 2020.
[16] T. V. Nguyen, "Exploring applications of big data analytics in supply chain management," Ph.D. dissertation, Univ. of Greenwich, London, U.K., 2019.
[17] C. S. Nwaimo, O. M. Oluoha, and O. Y. E. W. A. L. E. Oyedokun, "Big data analytics: Technologies, applications, and future prospects," Iconic Res. Eng. J., vol. 2, no. 11, pp. 411–419, 2019.
[18] C. M. Olszak and J. Zurada, "Big data in capturing business value," Inf. Syst. Manage., vol. 37, no. 3, pp. 240–254, 2020.
[19] K. M. Rahman, "A narrative literature review and e-commerce website research," arXiv preprint, arXiv:1806.07833, 2018.
[20] T. Rahat, N. Sultana, A. Mahmud, and A. Khanam, "AI for market sentiment and consumer behavior prediction," 2020.

[21] M. X. Wang and J. Q. Wang, "New online recommendation approach based on unbalanced linguistic label with integrated cloud," Kybernetes, vol. 47, no. 7, pp. 1325–1347, 2018.
[22] P. Zhao, Z. Ding, M. Wang, and R. Cao, "Behavior analysis for electronic commerce trading systems: A survey," IEEE Access, vol. 7, pp. 108703–108728, 2019.
[23] X. Zhao, L. Xia, J. Tang, and D. Yin, "Deep reinforcement learning for search, recommendation, and online advertising: A survey," ACM SIGWEB Newslett., Spring 2019, pp. 1–15.